

ND-A186 884

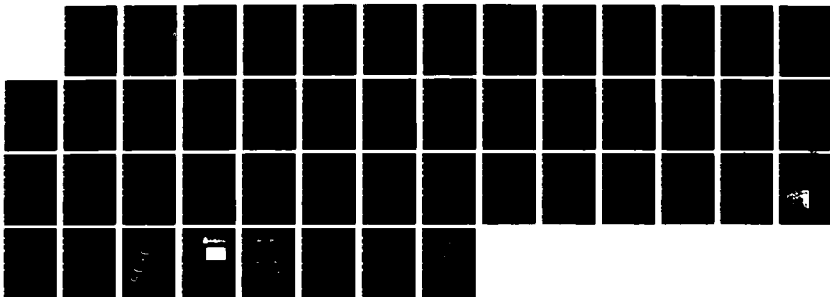
THE ROLE OF PHOTODETECTORS IN OPTICAL SIGNAL PROCESSING
(U) DUKE UNIV DURHAM NC G W ANDERSON ET AL 13 OCT 87
ARO-23620 1-PH DAAL03-86-C-0006

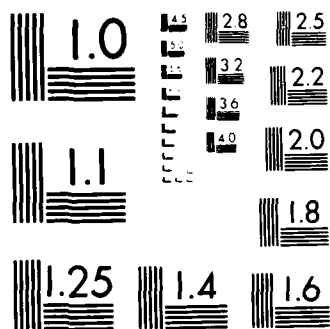
1/1

UNCLASSIFIED

F/G 17/5

NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(2)

AD-A186 804

TITLE

The Role of Photodetectors in Optical Signal Processing

TYPE OF REPORT (TECHNICAL, FINAL, ETC.)

Final Report

AUTHOR (S)

G. W. Anderson A. VanderLugt
B. D. Guenther
J. A. Hyncek
R. J. Keyes

DATE

October 13, 1987

DTIC
ELECTE
DEC 16 1987
S D
CD

U. S. ARMY RESEARCH OFFICE

CONTRACT / GRANT NUMBER

DAAL03-86-C-0006

INSTITUTION

Duke University

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED.

17 12 9 002

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION		1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION STATEMENT OF REPORT	
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE		4. PERFORMING ORGANIZATION REPORT NUMBER(S)	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)		5. MONITORING ORGANIZATION REPORT NUMBER(S)	
6a. NAME OF PERFORMING ORGANIZATION Duke University		6b. OFFICE SYMBOL (If applicable) SLCRO-PH	
7a. NAME OF MONITORING ORGANIZATION U. S. Army Research Office		7b. ADDRESS (City, State, and ZIP Code) P. O. Box 12211 Research Triangle Park, NC 27709-2211	
8a. NAME OF FUNDING / SPONSORING ORGANIZATION U. S. Army Research Office		8b. OFFICE SYMBOL (If applicable) SLCRO-PH	
9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER DAAL03-86-C-0006		10. SOURCE OF FUNDING NUMBERS	
10. SOURCE OF FUNDING NUMBERS		PROGRAM ELEMENT NO	
PROJECT NO		TASK NO	
WORK UNIT ACCESSION NO			
11. TITLE (Include Security Classification) The Role of Photodetectors in Optical Signal Processing			
12. PERSONAL AUTHOR(S) G.W. Anderson, B.D. Guenther, J.A. Hyncek, R.J. Keyes, A. Vanderlugt			
13a. TYPE OF REPORT Final		13b. TIME COVERED FROM 11-4-85 TO 11-3-86	
14. DATE OF REPORT (Year, Month, Day) 10-13-87		15. PAGE COUNT 45	
16. SUPPLEMENTARY NOTATION The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.			
17. COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB-GROUP	
		Optical Processing, Detectors	
19. ABSTRACT (Continue on reverse if necessary and identify by block number) Optical signal processing applications place demands on photodetector arrays beyond those encountered in image sensing applications. We review the basic requirements and show that increased dynamic range and non-linear decision operations that lead to reduced output data rates are the key improvements needed for both one and two dimensional arrays. Arrays of high-speed photodetector elements, with integrated post-detection circuitry, are also needed. We suggest some possible methods for achieving these goals; our main object is to stimulate the photodetector community to design and fabricate more useful devices.			
20. DISTRIBUTION / AVAILABILITY OF ABSTRACT <input type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT <input type="checkbox"/> DTIC USERS		21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL B. D. Guenther		22b. TELEPHONE (Include Area Code) 919 549-0641	
		22c. OFFICE SYMBOL SLCRO-PH	

The Role of Photodetectors in Optical Signal Processing

by

G. W. Anderson¹

B. D. Guenther²

J.A. Hynecek³

R. J. Keyes⁴

A. VanderLugt⁵

ABSTRACT

Optical signal processing applications place demands on photodetector arrays beyond those encountered in image sensing applications. We review the basic requirements and show that increased dynamic range and non-linear decision operations that lead to reduced output data rates are the key improvements needed for both one and two dimensional arrays. Arrays of high-speed photodetector elements, with integrated post-detection circuitry, are also needed. We suggest some possible methods for achieving these goals; our main objective is to stimulate the photodetector community to design and fabricate more useful devices.

¹ G.W. Anderson is with Naval Research Lab, Code 6813, 4555 Overlook Ave., S.W. Washington, D.C. 20375-5000

² B.D. Guenther is with Army Research Office, Research Triangle Park, N.C., 21211

³ J.A. Hynecek is with Texas Instruments, P.O. Box 225012, 3500 Central Expressway, Dallas, Texas 75265

⁴ R.J. Keyes is with Lincoln Laboratories, P. O. Box 73, Lexington, Mass 02173

⁵ A. VanderLugt is with North Carolina State University, Electrical and Computer Engineering, Box 7911, Raleigh, N.C., 27695

ALL INFORMATION CONTAINED
HEREIN IS UNCLASSIFIED
DATE 10-10-80 BY 1045
A-1

1. Introduction

A key attribute of optical processing is that computationally intensive operations can be performed using the highly parallel structure of the processing architectures. With few exceptions, however, the input and output signals are functions of time; we therefore need devices that connect the time and space coordinates of optical systems with the time coordinate of electronic systems. Various input devices have been or are being developed for converting time signals into space/time signals. In this paper, we focus on the output devices needed to convert the optically processed signals into time signals that can be further processed.

Most parallel optical processing architectures require an array of photodetector elements at the output of the system. Some performance parameters of currently available imaging devices are entirely acceptable. For example, linear arrays having up to 4096 elements, integrated on a single chip, are now available; this number is adequate for many one-dimensional processing applications. In contrast we need 2000 x 2000 element area arrays, but current two-dimensional arrays are limited to order of 1000 x 1000 elements. The evolution toward high definition television may provide the desired number of elements in the near future. Throughout this paper we distinguish between performance characteristics that will be obtained through evolutionary developments and those that can be achieved only through revolutionary developments; the latter are of greatest interest and importance for optical signal processing.

In Section 2 we describe the nature of some basic optical processing systems to establish required performance parameters of photodetector arrays. In Section 3 we review the basic device physics and current performance levels. We compare the noise mechanism for single element devices with those of arrays and show that we are not always quantum noise limited. In Section 4 we describe some characteristics desired of future devices, with particular attention to increasing the dynamic range and reducing the readout rates.

2. Background on Optical Data Processing

We provide an introduction to optical processing, with particular emphasis on the function of, and the requirements on, the photodetector component of the system. There are several review papers that cover the subject in greater detail^{1,2}, and journal issues have been devoted to special areas of optical

processing^{3,4,5,6}. Optical signal processing has its roots in the work of Abbe who showed, at the turn of the century, how an image formed by a coherently illuminated microscope could be modified by altering the diffraction pattern produced by an object. This diffraction pattern, more generally called the spatial frequency distribution, is the Fourier transform of the object; this transform plays a central role in optical processing.

Until 1960, the outputs of optical processing systems were generally detected, recorded, and displayed using photographic film; we might say that the first two-dimensional "photodetector array" was photographic film. Film requires a certain number of photons in a given time period, is sensitive to light in certain spectral ranges, has a dynamic transfer function (the H-D curve) and modulation transfer function, and has a given noise floor (often expressed in terms of granularity, grain size, or Selwyn's number). Although the terminology is different, the concepts are similar to those associated with photodetector arrays.

A key advantage of photographic film is that it has very high spatial resolution which can be chosen to match that of any realizable optical system. The major disadvantage is that film must be developed; as a result, it cannot operate in real-time with an electronic post-processing system. Real-time devices such as vidicon or orthicon image tubes can be used as the output detector but these devices have severe limitations such as limited dynamic range and geometric fidelity. The next improvement was the development of one- and two-dimensional photodetector arrays based on photo-sensitive CCD structures. These devices provide a new flexibility of operation and have many desirable features. They too, however, were developed primarily for imaging applications such as electronic newsgathering, surveillance, scanning, or inspection. As a result, these devices are not, in general, tailored for use in optical processing applications.

We now describe some basic forms of optical processing systems; the algorithms we implement are generally related to either spectrum analysis or correlation. These examples are selected to identify those features of optical processing that place certain demands on the photodetectors.

2.1 Spectrum Analyzers

Spectrum analysis is probably the most widely used algorithm in the physical sciences for gaining information about an unknown signal. Consider the optical system shown in Figure 1a. An image, recorded on a transparency, is placed in

plane P_1 and is illuminated by coherent light derived from a point source, at wavelength λ , via the lens L_1 . A key feature of this system is that the complex valued light distribution in plane P_2 , the back focal plane of lens L_2 , is the Fourier transform of the light distribution in plane P_1 :⁷

$$S(\alpha, \beta) = \int_{-\infty}^{\infty} a(x, y) s(x, y) \exp[j2\pi(\alpha x + \beta y)] dx dy \quad (1)$$

where $s(x, y)$ is amplitude of the signal, $a(x, y)$ is the aperture function which provides weighting to control sidelobe levels and establishes the boundaries of the signal to be processed, x and y are the spatial coordinates of plane P_1 , and α and β are the spatial frequency variables associated with the coordinates of plane P_2 . Since all planes in an optical system have spatial coordinates, we relate the spatial frequency variable to the coordinates ξ and η in plane P_2 by

$$\alpha = \frac{\xi}{\lambda F_2} \quad (2)$$

$$\beta = \frac{\eta}{\lambda F_2} ,$$

where F_2 is the focal length of lens L_2 .

If $s(x, y)$ contains some regular features such as the street pattern of a city, the spectrum $S(\alpha, \beta)$ will show strong spectral content in orthogonal directions. The spacing and the width of the sidelobes in the most intense part of this spectrum will indicate the period of the street spacings. In contrast, the spectrum associated with natural terrain is generally more uniformly distributed at all frequencies, with no predominant peaks. As an illustration of two-dimensional spectrum analysis, we show, in Figure 1b, a photograph of the ocean surrounding one of the Caribbean islands. Images of oceans produce spectra that give clues as to wind direction and ocean depth. From the spectra of subregion of the sea shown in the inserts, we note that strong diffraction of the light occurs in directions normal to the wave patterns. Where there are two interfering wave patterns, two relatively strong diffraction patterns are present. Where the surface waves, due to a shift in the wind direction, are propagating in a direction different from that of the major swells, the higher spatial frequencies have a direction differing from the lower frequencies. If we arrange for small portions of the input to be illuminated in a scanning mode, the spectra of these

subregions could be detected to give an indication of the sea state and its variation from one region to another.

2.1.1 A Special Photodetector Array

A photodetector array for detecting these spectral features consists of a set of wedge and annular photosensitive areas^{8,9,10}, as shown in Figure 1c. In this device, one half of the area contains N photoconductive surfaces in the shape of wedges. The output signals $\Omega_1, \Omega_2, \dots, \Omega_N$ indicate the degree to which $s(x,y)$ has spectral content at specific angles. The other half of the detector consists of $N+1$ photodetector areas in the shape of rings. The output signals $\rho_0, \rho_1, \rho_2, \dots, \rho_N$ indicate the relative energy present in $s(x,y)$ at various spatial frequency bins. Since $S(\alpha,\beta)$ is symmetrical about the origin, no information is lost in the detection process other than that which falls between the active areas. The object can be classified by processing the Ω and ρ values according to algorithms developed for specific applications. The details of the postprocessing is not important here; the requirements on the photodetector are. We now consider some of these requirements.

2.1.2 Spectral Sensitivity

The operating wavelength is dictated by the diffraction efficiency of the input modulator, scattering in the optical system, and the availability of compact, efficient sources. The normal source is a laser which provides coherent light; the photodetector should have a high sensitivity at the wavelength generated by the laser but need not have a broad spectral response. The most powerful lasers used in optical processing are water-cooled Argon Ion lasers with strong spectral lines at 488.0 nm and 514.5 nm; power outputs are in the 0.2 to 20 W range. Helium-Neon lasers, emitting at 632.8 nm are more often used; these lasers are air-cooled, compact, and reliable. They are generally less powerful, however, with typical outputs in the 10-50 mW range.

In recent years, the trend is toward the use of semiconductor lasers such as injection laser diodes (ILD) which emit light around 830 nm and produce 10-40mW of power. More powerful ILD's having 200mW of output power have recently been announced and even more powerful ones are being developed. For sake of illustration throughout this paper, we take $\lambda = 632.8$ nm and let the laser power be $P_0 = 10$ mW; these figures should not lead us too far astray in our calculations.

2.1.3 The Number of Elements

An important class of spectrum analyzers are those that use acousto-optic devices to convert time signals into space/time signals suitable for real-time processing. Acousto-optic devices, often called Bragg cells, consist of an interaction media such as glass or crystal to which a transducer is bonded. The time signal $s(t)$ drives the transducer so that pressure waves are launched into the interaction medium. The pressure wave induces a change in refractive index so that the light is modulated in both space and time by the traveling wave. Further details regarding the design parameters for Bragg cells can be found in the literature.^{11,12}

The cell will support a range of frequencies from f_1 to f_2 to provide an RF bandwidth of $W = f_2 - f_1$. The Bragg cell has a physical length L so that the time required for a signal to transit the cell is $T = L/v$, where v is the velocity of the acoustic wave within the cell; the time-bandwidth product of the cell is then simply TW . The time-bandwidth product establishes the number of elements required in the photodetector array; we must have $N = 2TW$ elements to avoid loss of information.

Common values for the TW of Bragg cells are of the order of 1000, which is almost independent of the interaction material or the bandwidth. For example, we may construct a Bragg cell having $W = 100$ MHz and $T = 10 \mu\text{sec}$ or one with $W = 1000$ MHz and $T = 1 \mu\text{sec}$. A somewhat higher TW can be obtained when tellurium dioxide (TeO_2) is used in a slow shear wave mode; in this instance we might have $W = 50$ MHz and $T = 40 \mu\text{sec}$ so that $TW = 2000$.

2.1.4 Array Geometry

Having established the required spectral sensitivity and number of elements, we now consider the question of the array geometry. The Bragg cell can be used to replace photographic film in a spectrum analyzer as shown in Figure 2a. Cylindrical lenses C_1 and C_2 are used to efficiently illuminate the Bragg cell; in effect, they cause the cell to have an apparent height equal to that of the cylindrical lenses. Lens L_2 , with focal length F_2 then produces, at plane P_2 , the Fourier transform of the information contained in the Bragg cell. Since there is no information in the vertical direction, all the light is focused onto the $\beta=0$ axis. The spectral content of the drive signal $s(t)$ is displayed in the horizontal direction, centered at a position ξ_c that represents a spatial frequency α_c corresponding to the temporal frequency f_c . By ignoring nonessential terms, we

have that

$$S(\alpha, t) = \int_{-\infty}^{\infty} a(x) s(t-x/v) e^{j2\pi\alpha x} dx \quad (3)$$

where $a(x)$ is the aperture function in plane P_1 . From (3) we see that the spectrum is a function of both the spatial frequency α and the current time t . In essence, $S(\alpha, t)$ represents a sliding window Fourier transform of $s(t)$, with $a(x)$ representing the window function.

The appropriate photodetector array for such a one-dimensional spectrum analyzer is a linear array as shown in Figure 2b. The array consists of N active elements (shown shaded), each of width d' and height h . The detector pitch is d so that $c = d'/d$ is the spatial duty cycle of the array elements. The angular separation between adjacent frequencies and the focal length of the transform lens determine the dimensions of the array. For a Bragg cell of length L , the angular separation between minimum resolvable frequencies is simply $\Delta\theta = \lambda/L$ so that the frequencies are spaced, at plane P_2 , a distance $\Delta\theta F_2$. We need at least two detector elements per frequency to satisfy the sampling theorem in the Fourier domain; we therefore conclude that

$$d = \frac{\lambda F_2^2}{2L} \quad (4)$$

In a well designed optical system the aberrations can be kept under control if the relative aperture, defined as $2L/F_2$, is less than $1/10$. Thus, an ideal pitch for the photodetector elements would be $d' \approx 10\lambda$, which is of the order of $6\mu\text{m}$. The pitch on currently available linear arrays is typically $12\mu\text{m}$; as a result, the length of the optical system must be increased beyond the ideal. Although this may not be a problem for ground based systems because the volume of the optical system is generally a small fraction of that of the electronics, the situation is reversed in airborne systems where the volume and weight of the spectrum analyzer must be minimized.

In the vertical direction we have a somewhat different criterion for setting the value of h . The Bragg cell performance is optimized when the illumination height equals that of the transducer, which is generally of the order of $100\mu\text{m}$. Since there is no sampling requirement in the vertical direction, the height of the photodetectors is set by the requirement that we collect the light

efficiently, consistent with minimizing the length of the optical system. For typical geometries, this requirement leads to a range of $h = 120 \mu\text{m}$, for the ideal case, to $h = 240 \mu\text{m}$, for the more typical case when $d = 12 \mu\text{m}$. The photodetector elements should therefore have aspect ratios of 10-20. The required geometric accuracy for most spectrum analysis and correlation applications is that center spacings be within $\pm 1\%$ of nominal, with a cumulative error of $\pm d/10$ over the length of the array.

2.1.5 Readout Rate

We now consider the electronic format and readout rates for linear arrays. In applications such as spectrum analysis, $S(\alpha, t)$ may be a slowly varying function of time so that the sampling rate can be set fairly low without loss of information. Integration on the array, then, may range from a few milliseconds to a few seconds, provided that the detector is not saturated. If the required integration time is very long, as in a radiometer application, the contents of the array must be readout, digitized, stored, and accumulated in a post-detector memory whenever saturation is approached.

In other applications the spectrum may change rapidly so that we want to read the array once every T seconds to avoid loss of information. Unfortunately, the required temporal sampling rate cannot be sustained by the CCD transfer rates. As an example, suppose that we implement a spectrum analyzer having a 400 MHz bandwidth and an effective processing time of $T = 1.25 \mu\text{s}$ so that the time-bandwidth product is 500. We then need a 1000-element array to accurately sample the spectrum spatially. If we read the entire array in T seconds, the CCD output rate is $800(10^6)$ samples per second; this rate is too high for both the CCD and the subsequent digital electronic post-processing system to handle directly. A partial solution to this problem is to use multiple video lines to reduce the rate per line; the aggregate rate, however, remains unaffected.

We need "smart" arrays in which some of the post-processing functions are included on the photodetector chip. If we could implement some logic functions at the element level, the transfer data rates and the complexity of the subsequent electronics would be drastically reduced. For example, suppose that we transfer information only if a photodetector element exceeds some pre-selected threshold. Or, suppose that we transfer information only if the instantaneous intensity exceeds some pre-set value. These operations require that we build some associated circuitry for each element or set of elements in the array. The implications of

developing such a capability are enormous for many applications because the transfer rates associated with processing a very wideband received signal can then be reduced by factors of 10^2 to 10^3 or even more.

2.1.6 Dynamic Range

A key requirement of arrays is a large dynamic range; we often need to achieve a 70-80 dB dynamic range; the dynamic range is always referenced to the signals applied to the Bragg cell. For example, when the cell is driven so that the system is linear in light amplitudes, the Fourier transform $S(\alpha, t)$ is an amplitude function. The resultant light intensity and, in turn, the optical power is proportional to the square of the amplitude A which is proportional to the square of the applied voltage V . Thus, in a power spectrum analyzer, we have a direct correspondence between the optical output power and the electrical input power.

The current i_s produced by a photodetector is proportional to the detected optical power. In turn, the dynamic range of the photodetector is based on $10 \log(i_s^2)$ so that if the detector and its circuitry can sustain a dynamic range of say 60 dB, the received RF power can vary by only 30 dB; this arises because $i_s \propto V^2$ so that $i_s^2 \propto V^4$ in a direct detection system.

A significantly different situation arises when we use a heterodyne technique in which the detected optical power is $P_d \propto 2 \sqrt{P_r P_s}$, where P_r is the fixed reference beam power and P_s is the signal beam power. The current i_s from the detector is now proportional to $\sqrt{P_s}$ so that $i_s \propto V$; in turn, the output signal-to-noise ratio and dynamic range calculations are based on $i_s^2 \propto V^2$. As a result, a detector dynamic range of 60 dB is fully available to accommodate a 60 dB variation in the received RF signal. This heterodyne advantage can be obtained either spatially or temporally through optical interferometric schemes. The price we pay for spatial heterodyning is the need for more photodetector elements; the price we pay for temporal heterodyning is the need for an array of discrete elements that can be read out in parallel.

2.1.7 Sensitivity and Power Levels

Photometry is a science whose terminology is strange. In the photodetector literature, we have quantities such as lumens, lux, phots, stilbs, apostilbs, candelas, etc. and we ratio these by feet, square feet, steradians, centimeters, etc. to get even stranger quantities such as a nit (a candela per square meter). We shall use watts as the measure of power and millimeters as the unit of distance.

In a spectrum analyzer we must operate the Bragg cell at fairly low efficiencies to control the intermodulation products. A common solution for reducing intermodulation products to an acceptable level is to operate the cell at a diffraction efficiency of no more than 1% per frequency, with no more than 20 or so frequencies at saturation. The rest of the system (lenses, mirrors, beamshaping, etc.) may be about 30% efficient. Given a laser power of $P_0 = 10\text{mw}$, we find that the maximum power that a photodetector will intercept is of the order of $15\mu\text{W}$. The weakest signal we want to detect may be 60-70 dB below this level.

2.1.8 Blooming and Electrical Crosstalk

It is important that the array be readout before the elements saturate. However, if we want to detect weaker signals by means of longer integration times without reading the array, we must ensure that the excess charge is properly drained away so that spillover into adjacent elements does not mask the weak signals. In general, we want the crosstalk level to decrease at a rate of at least 10 dB per element away from a saturated element, to a level of at least -70 dB for all elements further away than the seventh element.

2.1.9 Linearity and Uniformity of Response

Single element photodiodes have a large linear dynamic range. At some intensity, however, saturation sets in and the slope of the dynamic transfer curve decreases. Since we often introduce a companding or compression scheme to facilitate the readout and display of the information, a high degree of linearity is not required. A more important characteristic is that the response is monotonic so that we can establish an inverse mapping that allows us to measure the input intensity to the required accuracy.

The saturation phenomena is slightly different for CCD arrays. The charge accumulates until the well is full; additional charge is not readout so that there is a discontinuity in the derivative of the transfer curve. Since a unique inverse mapping cannot be obtained for this case, saturation in CCD arrays must be avoided.

A uniformity of response across the elements of an array of $\pm 10\%$ is generally adequate; this degree of uniformity is easily met with current technology.

2.2 Two-dimensional Spectrum Analyzers

We can configure an optical system that uses a two-dimensional spatial light modulator to format wideband electronic signals; this modulator replaces the photographic film used in the system shown in Figure 1.^{13,14} More recently, the trend is to use orthogonal Bragg cells to provide the desired format.^{15,16} The

output of the system is a two-dimensional spectrum wherein coarse frequencies are displayed on the horizontal axis; each coarse frequency, in turn, is further decomposed by the action of the second Bragg cell to provide a fine frequency resolution in the vertical direction that is of the order of $1/T_0$, where T_0 is the total integration time. We generally need at least a 1000×1000 element photodetector array for this application; if we need to resolve a spatial carrier frequency to achieve a heterodyne detection, we need $\approx 2-4$ times as many elements in one of the directions.

This array must satisfy most of the performance measures of the linear array described in Section 2.1. The main difference is that the light intensity at the detector plane is significantly reduced. We suffer a 1000 fold reduction in light due to decomposing a single coarse frequency into 1000 fine frequencies in the vertical direction. The lower light level is compensated by using longer integration times which, in turn, leads to lower readout rates. For example, if we use 1/30 sec as our readout period, the sample rate is $30(10^6)$ samples per second. There are, however, applications where we would like to read the array at a rate of 1000-2000 frames per second so that rapidly changing signals such as frequency hoppers can be detected.

2.3 Heterodyne Spectrum Analyzer

One way to overcome the dynamic range limitation is to temporally heterodyne the complex spectrum with a coherent reference beam derived from the same light source^{17,18}. Figure 3a shows a heterodyne system in which the lower branch is the same as a power spectrum analyzer. In the Fourier domain the reference beam from the upper branch provides a distributed local oscillator that heterodynes all spectral components of the received signal to a fixed temporal frequency. The output of each photodetector is of the form $B + A_k \cos(2\pi f_0 t + \phi_k)$, where B is a bias term, A_k and ϕ_k are the Fourier coefficient and phase of the k^{th} frequency component, and f_0 is a conveniently chosen IF frequency. Each element in the array must therefore respond to a narrowband signal centered at f_0 . The low frequency terms, lumped into a bias term B , must be eliminated by filtering. As a result, we cannot use a conventional CCD photodetector structure; instead, we need a totally new kind of array in which we have parallel readout so that the signal from each element can be processed independently, as shown in Figure 3b. The basic advantage of the heterodyne approach is that the dynamic range is nearly doubled in dB's, so that a 30-35 dB dynamic range can be extended to 60-70 dB.

2.3 Correlation Applications

The physical demands on photodetector arrays, such as the number of elements, spacings, sensitivities, and so forth are much the same for correlation applications as they are for spectrum analysis. The dynamic range requirement can be relaxed because the peak signal intensity is generally set by factors such as the target size, laser power, and optical efficiencies. The detected signal strength, therefore, does not vary over a wide range as it does in spectrum analysis. There are, however, some new features of the detector array that are highly desirable.

In Figure 4 we show a classical two-dimensional system used for pattern recognition. It is an extension of that shown in Figure 1, with a spatial filter being placed in the Fourier domain P_2 and a photodetector array placed in the image plane P_4 of the input object; this output plane is sometimes called the correlation plane. The filter is constructed to match, in both amplitude and phase, the Fourier transform of a known signal. If this signal occurs in the received data placed in the input plane, a strong correlation peak results at the correlation plane at the image position of the signal. The photodetector must then detect the output light intensity and decide whether the correlation peak exceeds some preset threshold.

The photodetector array must contain as many elements as there are degrees of freedom in the photographic image. It is not uncommon to work with 125mm wide film having 100 cycles/mm resolution; to fully match this capacity we would need a square array having 25,000 elements on a side to satisfy the sampling theorem. There are no arrays that meet the desired number of elements. But correlation systems for pattern recognition have some features that may help to solve this problem if the proper nonlinearity can be incorporated into the array. Consider a frame of imagery that contains several targets as shown in Figure 5a. A matched filter will detect signals when its orientation and scale are correct and collapse most of the signal energy into a 2×2 pixel space as shown in Figure 5b. The ratio of the number of pixels covered by the target to the number covered by the correlation peak is of the order of 10^4 for typical targets. Since the targets generally do not overlap, we could partition the correlation plane into $4TW/10^4$ possible correlation regions and simply inquire as to which regions are active.

For the example given of 125mm film with 100 cycles/mm resolution, we then require a photodetector array of only 250×250 resolution elements which is reasonable. But these elements must be of a very special type. They must not integrate the light over their surfaces. Instead, they must respond in a highly

nonlinear fashion only when the intensity over some very small part of their surface exceeds threshold. This requirement is more easily understood by considering, as in Figure 5c, an intensity profile across one scan line of the correlation plane. Suppose that the ratio of peak correlation signal to mean square noise is of the order of 10^3 and that the number of pixels contained in the target is 10^4 . If we were to use a photodetector surface whose response is linear, the integrated background intensity would be roughly 10 times that of the contribution by the correlation peak because, although the background noise intensity is small, there are 10^4 contributing elements to the result. We would then have converted a +30dB SNR situation into a -10 dB signal-to-noise ratio condition; this is clearly unacceptable.

The desired transfer characteristic for each photodetector is shown in Figure 5d; when the light intensity is below a given value at a particular position, that element does not contribute to the output. When the intensity threshold is exceeded due to the presence of a correlation peak, the output goes to its maximum value as desired. Only a few of the 100x100 possible positions in any one of the 250x250 set of elements is therefore "punctured", the remaining ones being below the breakdown voltage. If we had such a device for pattern recognition, we could very easily conceive of processing high quality imagery--an application that we have basically ignored to date.

2.4 Single Element Devices

Finally, there are optical processing applications wherein a wideband single element photodetector is needed. The format is preferably rectangular with a high aspect ratio (e.g., a detector 4mm long by 0.40mm high.) Bandwidths up to 1 GHz are required and the key is keeping the capacitance of the device under control while maintaining the bandwidth.

2.5 Summary of Needs

We have briefly touched on various optical processing architectures that illustrate some of the requirements on photodetectors. There are, of course, many other applications such as adaptive filtering or ambiguity surface and Wigner distribution generation; these applications have requirements similar to those given above. Photodetector arrays are also needed in applications such as synthetic aperture radar processing or tomographic processing. We have defined five generic types of detectors: (1) a wideband single element detector, (2) a one-dimensional array of discrete (parallel readout) detectors, (3) a one-dimensional array of

integrating (serial readout) detectors, (4) a two-dimensional array of discrete detectors, and (5) a two-dimensional array of integrating detectors. We can summarize these needs as follows:

One-dimensional Arrays

- o At least 2096 elements in (CCD format)
- o At least 256 elements (discrete, wideband)
- o Multiple video lines (total rate of > 100 MHz)
- o Wide dynamic range, low crosstalk, high linearity, low fixed pattern noise, low reflections, not polarization dependent
- o Desire either serial or random access read out of elements
- o Temporal change detection from frame-to-frame
- o Spatial rate of change detection over subregion of the array
- o On-chip processing to calculate centroids, neighborhood operations, etc
- o Bandwidths per element range from <1 MHz (serial readout) to >400 MHz (discrete, fully parallel readout)
- o High sensitivity, particularly at light wavelengths of ≈ 600 - 850 nm

Two-dimensional Arrays

- o Need 2000×2000 element arrays
- o Other specialized array formats:
 - 3 x 2096 for direction-of-arrival spectrum analysis
 - 3 x 3 discrete for wideband signal acquisition with Doppler
 - 25 x 25 for terminal guidance
 - 250 x 250 for pattern recognition (thresholding and latching)
- o AC coupled, lock-in amplifier effect
- o Other requirements are as for one-dimensional arrays

3 Device Characteristics

A photodetector converts optical energy to electrical energy. Four basic types of detectors can be used in an optical processor; these are two-dimensional vacuum tube sensors, single element solid state sensors, one-dimensional solid state detector arrays, and two-dimensional solid state detector arrays.

Vacuum tube detectors have advantages such as operation in a photon noise

limited regime; however, we shall focus on the use of solid state devices because of their small size, high reliability, low voltage requirements and the possibility of integrating the detector with signal processing electronics.

There are three basic classes of solid state photodetectors: photoconductors, depletion layer detectors and avalanche photodetectors

3.1 Elementary Device Physics

The major device physics parameters which are critical to acousto-optic signal processing photodetectors are quantum efficiency, responsivity, pixel-to-pixel crosstalk, charge capacity, intrinsic detector transfer characteristics, detector dynamic range, and noise. These device parameters are important because of the high dynamic range (≈ 80 dB or 10^8 in terms of photon flux) and high speeds required at the output of an optical processor. They have been treated in detail elsewhere^{19,20,21} and consequently will be covered only briefly here.

3.1.1 Photoconductors

A photoconductor is a junctionless semiconductor device whose conductivity increases when exposed to light. Incident photons with energy greater than the semiconductor bandgap produces electron-hole pairs which contribute to the device's conductivity until recombination takes place. The change in current density due to illumination is

$$\Delta J = q\mu\eta\phi\tau, \quad (5)$$

where q is the charge, μ is the mobility of the charge carrier, η is the quantum efficiency, ϕ is the light flux, and τ is the lifetime of the charge carrier. The current gain in a dc biased photoconductor given by

$$G = \frac{\tau}{t_0}, \quad (6)$$

where t_0 is the transit time of the faster charge carrier (usually the electron) across the active detector region; gains of 10^3 or more can be obtained in silicon. For $\tau \gg t_0$, the bandwidth of the device is established by the carrier lifetime:

$$B = \frac{1}{\tau}. \quad (7)$$

The gain-bandwidth product is given by (6) and (7):

$$GB = \frac{1}{t_0}, \quad (8)$$

and it is a constant for a given material and detector design. For silicon, the value of GB is between 1 and 10^3 MHz.

The primary noise source in the photoconductor is Johnson noise, due to the bias current, which limits applications to the detection of light signals with limited dynamic range. New photoconductor designs such as Photo-FETS and devices with very low doped materials show promise of overcoming these disadvantages. At present, however, photodiodes are the most widely used devices.

3.1.2 Photodiodes

The depletion layer photodiode is a reversed biased pn junction or Schottky barrier diode whose reverse current is modulated by charge carriers produced by photons in or near the depletion layer of the device. A typical device is shown in Figure 6. A pn junction is reversed biased, creating an electric field of the order of 10^4 v/cm at the junction. This field depletes a region, of width W , of all carriers; this region is called the depletion region. When a photon with an energy greater than the bandgap is absorbed in the depletion region, an electron and hole are created. The electric field causes the carriers to rapidly drift to the electrodes producing a photocurrent. It is possible that all of the photons will not be absorbed in the depletion region but will penetrate to a much greater distance into the n-doped material of Figure 6. The holes (minority carriers) produced in a region of width L are able to diffuse to the electrode before they recombine. These holes also contribute to the photocurrent. If α is the absorption coefficient, then the photon flux at a distance x into the n-doped material is

$$\phi = \phi_0 e^{-\alpha x}, \quad (9)$$

where

$$\phi_0 = \frac{P_0}{h\nu} \quad (10)$$

is the number of incident photons in a beam with an optical power of P_0 and a frequency of ν . Any photons existing beyond a distance $x=W+L$ do not contribute to the photocurrent and are lost; we have also ignored absorption in the electrical contact layer.

The quantum efficiency is defined as the number of charge carriers produced by the detector divided by the number of photons incident on the detector

$$\eta = \left[\frac{\phi_0 - \phi(W+L)}{\phi_0} \right] = (1-r) \left[1 - e^{-\alpha(W+L)} \right] = (1-r) \left[1 - \frac{e^{-\alpha W}}{1+\alpha L} \right], \quad (11)$$

where r is the Fresnel reflection coefficient at the surface of the detector. Reduction of reflection losses at the external surface of the detector through the use of an antireflection coating will increase the quantum efficiency by increasing the number of photons reaching the depletion layer. The photocurrent density generated by a detector is

$$J = q\phi_0\eta \approx q\phi_0 \left[1 - \frac{e^{-\alpha W}}{1+\alpha L} \right] (1-r). \quad (12)$$

To maximize the photocurrent it is obvious that αW and αL should be as large as possible. The depletion layer width is inversely proportional to the doping concentration:

$$W = \sqrt{\frac{2\epsilon(V_a + V_c)}{q} \left(\frac{1}{N_a} + \frac{1}{N_d} \right)}. \quad (13)$$

where V_c is the contact potential, V_a is the applied voltage, V_c is the contact voltage, N_a is the acceptor concentration and N_d is the donor concentration. The drift component of the current is a fast transport mechanism, as short as 0.1 nsec for $W=10\mu\text{m}$ in silicon, while the diffusion component is relatively slow, 40 nsec in silicon for $L=10\mu\text{m}$. Thus for high speeds, L should be kept small and W should be large. The junction capacitance, which affects the speed of the diode through the RC time constant of the output circuit, is given by

$$C = \frac{\epsilon A}{W} = \frac{A}{\sqrt{\frac{2}{q\epsilon}(V_a + V_c) \left(\frac{1}{N_a} + \frac{1}{N_d} \right)}}, \quad (14)$$

where A is the diode junction area. As can be seen, low doping and large applied fields increase W and thereby increase the frequency response of the detector. For silicon we have $V_c=0.6\text{V}$ and for GaAs we have $V_c=1.1\text{V}$. An intrinsic material would be used to maximize the value of W because there are few donors or acceptors and N_a and N_d are nearly zero. The width L is a function of the minority carrier lifetime and the diffusion rate of minority carriers.

The absorption coefficient consists of three coefficients. Only the interband absorption coefficient α_B contributes to the quantum efficiency. The

other two coefficients are the free carrier absorption, α_C , and the scattering loss, α_S . The processes that give rise to these coefficients remove photons from the incident radiation without generating current. The quantum efficiency is thus reduced by the factor

$$\frac{\alpha_B}{\alpha_B + \alpha_C + \alpha_S} \quad (15)$$

The photodetector's performance is also a function of the wavelength of the illumination. The absorption coefficient is a function of wavelength and wavelengths much longer than the bandgap, E_g , will be associated with a small α_B . This means that the quantum efficiency will drop as the illuminating wavelength shifts to the long wavelength side of the bandgap due to deep penetration of photons into the substrate. The bandgap of the semiconductor should not be much smaller than the illuminating wavelength's energy. If it is, an excessive number of thermally generated charge carriers will be produced, resulting in a large dark current, I_d . The p^+ layer of Figure 6 introduces a short wavelength cutoff by absorbing photons near the surface. This limitation can be overcome by illuminating the device in Figure 6 from below. A second technique for overcoming the short wavelength limit imposed by the p^+ layer is to replace the layer by a metal film, producing a Schottky barrier photodiode whose operation is essentially the same as the p^+n junction. In an optical processing application, the only restrictions on operating wavelength placed on the system are the availability of a source and any restrictions placed on the system by the input and output devices.

To reduce the series resistance, and thus some of the noise, the p region is heavily doped. A heavily doped n region is added to the back of the intrinsic layer to serve as a second contact. When intrinsic material is used the result, shown in Figure 6, is a PIN diode. The high purity of the intrinsic layer allows a large depletion width to be created at low voltages which minimizes the noise current.

3.1.3 Avalanche Photodiodes

If the reverse bias voltage is set near the avalanche breakdown point (≈ 100 - 400 V), the gain of the detector can be raised above one. A region of high field of the order of 10^5 v/cm accelerates the charges to sufficient energy to create new carriers through impact ionization. The gain of this device is defined as the ratio of the multiplied current to the photon induced current

$$G \propto e^{(\beta_n + \beta_p)W}, \quad (16)$$

where β_n and β_p are the ionization rates for the negative and positive charge carriers respectively. The gain can also be written as a function of device parameters; the maximum gain is given by

$$G_{\max} \propto \sqrt{\frac{V_b}{nIR}}, \quad (17)$$

where V_b is the breakdown voltage, n is an experimentally determined parameter, i is the saturation current, and R is the series resistance of the diode (including space charge effects). The avalanche gain is a function of the bias voltage, V_b :

$$G \propto \frac{1}{1 - \left[\frac{V_b - IR}{V_b} \right]^n}, \quad (18)$$

and this dependence upon bias can lead to nonlinearities when a large dynamic range is required. The breakdown voltage is a function of the temperature and thus the avalanche gain varies with temperature. These effects require control of high bias voltages. Breakdown can occur at lower voltages due to edge breakdown or microplasma generation at localized defects. These effects can limit the gain that can be produced in a device.

When $\beta_p \ll \beta_n$, so that only the electrons contribute to the ionization process, the bandwidth is maximized and excess noise produced by the multiplication process is minimized. The relative size of the two ionization rates is characterized by the ratio

$$k = \frac{\beta_p}{\beta_n}. \quad (19)$$

As k approaches one, the gain becomes a strong function of the applied electric field which makes it difficult to manufacture devices with uniform characteristics. Nonuniform materials increase the randomness of the collision ionization, leading to an increase in the noise in the detector. The excess noise term is of the form

$$G^x \propto kG + (1 - k) \left[2 - \frac{1}{G} \right]. \quad (20)$$

If $k = 1$, the excess noise term is equal to G and $x = 1$. If $k = 0$, the excess noise term is a factor somewhere between 1 and 2. Typical k values of 0.02 in silicon yield an excess noise of 4 at a gain of 100. Values of k a factor of 2.5 times smaller have been reported.²⁰ Other noise processes and bandwidth limitations are the same as the limits of a PIN diode. There is a reduction in bandwidth with gain because the avalanche process takes time to develop.

A number of problems are associated with avalanche photodiodes. The major problem is the production of a device with a dislocation free substrate to raise the effective breakdown voltage level, and a doping concentration controlled to 0.1% to reduce the effects of random fluctuations. A second problem is obtaining enough surface passivation to reduce edge breakdown. The passivation also reduces surface leakage change with age which is a major cause of reliability problems. Arrays of these devices are more difficult to produce because of the need for guard ring structures to prevent edge breakdown.

3.2 Arrays

The detectors discussed above must be assembled into either one- or two-dimensional arrays before they can be used in an optical processing application. The detector arrays may be instantaneous or integrating devices whose outputs are addressed in parallel or in serial output lines, respectively. The detector array devices may be photoconductor arrays, charge coupled device (CCD) arrays, avalanche photodiodes (APDs), PIN diodes, or Schottky barrier diodes. If the arrays are nonintegrating, the signal from each detector is amplified by a single amplifier circuit which may be a logarithmic amplifier in order to provide some companding to handle a high dynamic range. If the arrays are integrating devices, they may be addressed by MIS shift registers or by CCD parallel-in, serial-out shift registers.

Two special problems arise when detectors are assembled into arrays. One is optical in origin and is called pixel-to-pixel crosstalk. The second is electronic in origin and is called fixed pattern noise. Photon produced electrons generated in one detector element and collected in another cause pixel-to-pixel crosstalk. The techniques used to isolate individual pixels affect the quantum efficiency. Fixed pattern noise arises from variations in the performance parameters of the array elements.

3.2.1 Pixel-to-Pixel Crosstalk

In Figure 7 the photodiodes are isolated by a region of opposite carrier concentration, here denoted as a n region, often called the tub. The thickness of the tub region below the depleted layer is important in limiting pixel-to-pixel crosstalk due to both thermally and photon generated carriers. Carriers formed due to absorption of photons below the tub layer do not contribute to crosstalk because they are not collected by any photodiode.

The quantum efficiency is affected by the techniques used to create the individual pixels. All photons absorbed between the pixels can contribute to the quantum efficiency but would also contribute to optical crosstalk. If grooves are etched between the pixels, however, or if channel stops formed by diffusion or ion implantation exist, as is commonly the case in CCD imagers, the pixel-to-pixel isolation is improved but the quantum efficiency is decreased. A deeper depletion region will result in collection of more charge, overcoming the loss associated with isolation but will increase the crosstalk arising from carriers produced deep in the device. A tradeoff must be made between crosstalk and quantum efficiency.

In the tub region the carriers produced travel by diffusion and thus affect the temporal performance of the detector. If the integration period is of the order of a microsecond, holdover or temporal crosstalk can occur. One method of solving this problem is to use a collection region. The collection region/substrate junction is important for limiting pixel-to-pixel crosstalk as well as temporal crosstalk.

A final source of noise in arrays is blooming. When the incident light is very high, blooming may be caused by charge spilling from one pixel to another. A number of methods of handling blooming have been developed such as adding a drain to the sensor structure to remove excess charge.²²

3.2.2 Fixed Pattern Noise

Fixed pattern noise occurs because of nonuniformities in gain or the dc offset in different detector elements. Leakage of signals from switching circuits also contribute to the fixed pattern noise. The dynamic range and measurement accuracy is reduced by the pixel-to-pixel nonuniformities.

4. Discussion of Representative Architectures

To achieve the high performance required of photodetectors, we examine the limits encountered in constructing such detectors and the discovery of new approaches that could improve their performance. The device structures are usually divided into two categories. First is that of nonintegrating detectors, where the photon flux is continuously converted into an electrical output signal. The second is that of integrating detectors, where the charge produced by the incident photons is first integrated and then converted into the output signal. The latter category can be further subdivided into two subgroups. One subgroup uses a feedback amplifier connected to the detector. The second subgroup uses an array of detectors such as in a CCD or CID device; the optical signal is converted into a charge packet, stored in a potential well, and is later transferred into a charge-to-voltage converting structure common to a group of detectors or to the whole array. To understand the fundamental limits of operation, functional differences, and common features of each of these detectors, we briefly describe their principles of operation.

The nonintegrating detector usually consists of a photodiode connected to a feedback amplifier as shown in Figure 8. This device has a wide dynamic range due, in part, to the feedback loop which maintains a constant bias across the photodiode, irrespective of the magnitude of the photocurrent that is generated in the detector. The dynamic range of this arrangement is the ratio of the maximum voltage which can be permitted across the feedback resistor, before reaching saturation, to the minimum noise floor.

The minimum noise floor consists of thermal noise and shot noise caused by the average current in the circuit. The shot noise term is

$$SN = 2e(I_s + I_d + I_h)B_n R_L G^m, \quad (21)$$

where e is the charge of an electron, B_n is the noise bandwidth, R_L is the load resistance, $G^m = G^{2+x}$ is a gain factor characteristic of avalanche photodetectors, and G^x is the excess noise introduced earlier. Typical values of m are 2.3-2.5 for silicon devices and 2.7-3.0 for III-V alloy devices. The average currents are I_s due to the signal and I_d due to the dark current. We also include a current I_h generated if we add a local oscillator to the signal to operate in a heterodyne detection mode. When we operate in a direct detection mode, the current I_h is equal to zero. The thermal noise is

$$TN = 4kTB_n, \quad (22)$$

where k is Boltzmann's constant, and T is the temperature in degrees Kelvin. The signal to noise ratio is then

$$SNR = \frac{\langle i_s^2 \rangle G^2 R_L}{2e(I_s + I_d + I_h)B_n R_L G^m + 4kTB_n} \quad (23)$$

where $\langle i_s^2 \rangle$ is the time average of the current produced by the signal.

The value of the load resistance is determined by the upper cutoff frequency f_c required to pass the signal:

$$R_L = \frac{1}{2\pi C_d f_c}, \quad (24)$$

We substitute (24) into (23) and rearrange terms to obtain

$$SNR = \frac{\langle i_s^2 \rangle G^2}{2e(I_s + I_d + I_h)B_n G^m + 8\pi k T C_d B_n f_c} \quad (25)$$

A few comments regarding (25) are in order so that we understand how the signal-to-noise ratio is to be calculated under the wide range of photodetector applications described in Section 2. First, we consider the class of nonintegrating devices. In some cases the signal may be at baseband so that

$B_n \approx f_c$; in other cases the signal may be narrowband so that $B_n \ll f_c$. In either event, we want to select a photodetector gain that maximizes the dynamic range, using the minimum laser power. This occurs when

$$G^m \approx \frac{4\pi k T C_d f_c}{e(I_s + I_d + I_h)}, \quad (26)$$

which reveals that avalanche diodes are most useful when the signal is wideband (large f_c) or when we use direct detection ($I_h = 0$). From (25) we see that, since $m > 2$, any gain larger than that indicated by (26) will result in a deterioration of the dynamic range. In heterodyne applications the local oscillator is usually sufficiently strong so that shot noise dominates the thermal noise (large I_h); avalanche detectors then give way to PIN detectors, for which $G = 1$.

Second, consider integrating photodetector array circuits which usually have low bandwidths so that the shot noise dominates the thermal noise and avalanche operation is of no advantage. We can then simplify the signal-to-noise

ratio expression to

$$\text{SNR} = \frac{\langle i_s^2 \rangle}{2e(I_s + I_d + I_h)B_n} \quad (27)$$

The minimum noise occurs when $I_s \approx I_d + I_h$. For direct detection applications, we see that the system performance is dark current noise limited; for heterodyne detection the performance is local oscillator current noise limited.

The dynamic range is the ratio of the maximum signal power to the minimum noise power. In a nonintegrating application, the maximum signal power is determined by the available laser power or by an allowable departure from linearity as saturation is approached. In an integrating application, the maximum signal power is set by the saturation level as determined by the charge capacity of the diode. Since the dynamic range is an important parameter in optical processing, we give the results for the four principal modes of operation:

DIRECT DETECTION; INTEGRATING

$$\text{DR} = \frac{\langle i_s^2 \rangle}{2eI_d c_d B_n} \quad (28a)$$

HETERODYNE DETECTION; INTEGRATING

$$\text{DR} = \frac{\langle i_s^2 \rangle}{4e(I_d + I_h) c_d B_n} \quad (28b)$$

DIRECT OR HETERODYNE DETECTION; NONINTEGRATING

$$\text{DR} = \frac{\langle i_s^2 \rangle G^2}{16\pi k T c_d B_n f_c} \quad (28c)$$

From these results, we might conclude that the dynamic range does not increase when heterodyne detection is used, contrary to our earlier claims. The advantage of heterodyne detection, which results in a doubling of the dynamic range in dB's, becomes apparent when we realize that the photodetector current in heterodyne detectors is linearly related to the optical signal amplitude instead of to the optical signal intensity as is the case for direct detection.

Suppose, for example, that P_h is the optical power at the detector from the local oscillator and that P_s is the optical power from the signal. The induced photocurrent is then

$$i_s = S \left[P_h + P_s + \sqrt{2P_h P_s} \cos(2\pi f_0 t) \right], \quad (29)$$

where S is the sensitivity of the photodetector in amps/watt, and where we have introduced a temporal IF frequency f_0 to help separate the desired third term from the first two bias terms. The average currents due to the local oscillator and to the signal are $I_h = SP_h$ and $I_s = SP_s$. These terms contribute to the shot noise; they do not appear in the numerator of our signal-to-noise ratio expressions because they are typically eliminated by a bandpass filter. After filtering we see that signal current is proportional to $\sqrt{P_s}$ so that the system is linear in optical signal amplitudes. Without the heterodyne action provided by the local oscillator, we see that the system is linear in optical signal intensities.

From these results, we see that the signal-to-noise ratio is always smaller than the dynamic range in a well designed system, a direct result of the Poisson's statistics of the noise in these photon detecting systems. In devices such as CCD's, for example, where the charge is generated electrically by filling a potential well, the noise in each signal charge packet is approximately constant and independent of the number of input electrons. It is, therefore, useful to extend the dynamic range by using a nonlinear element in the feedback of the amplifier without sacrificing the signal-to-noise ratio. Equation (23) remains approximately valid in this case since it does not directly contain the feedback element parameters.

The possibility of extending the dynamic range has been recognized by researchers who successfully designed and tested systems containing logarithmic amplifiers or piecewise linear amplifiers.²² The dynamic range obtained approaches 80 dB in power units which is satisfactory for optical signal processing applications. One problem encountered is a variation of the bandwidth with the signal level, as suggested by (24). This problem could, in principle be eliminated if the capacitance c_d is made variable so that the $c_d R_L$ product is kept constant. It is not difficult to incorporate a voltage variable capacitor (varicap) into the feedback loop, which changes its value by an order of magnitude. It may not be practical, however, to compensate exactly the variations in R_L by variations

in c_d , and some excess bandwidth must be designed into the system. This will, unfortunately, increase the noise floor of the amplifier somewhat.

Integrating detectors which use an amplifier with feedback will have a comparable performance. The operation of such a detector-amplifier system can be understood from the circuit given in Figure 9. The signal is integrated on the feedback capacitor c_d which is reset after the readout is completed. The signal-to-noise ratio and the dynamic range are expressed as before, except that the noise bandwidth is now defined as $B_n = 1/(2T_F)$, where T_F is the time interval between successive readouts of the diodes (the integration time). By analogy with the technique just described, we can also extend the dynamic range by using a nonlinear element in the feedback loop. The varicap could be easily incorporated into an integrated detector-amplifier array. An advantage of this approach is the flexibility in selecting of the integration time, which can be made variable to adjust to the average signal level and to always use the full dynamic range of the amplifier. An additional advantage available in integrating systems is the possibility of reducing the effects of the thermal noise. This can be achieved by a well known correlated double sampling signal processing method²³. In this approach a difference of two signal readings is formed between the outputs just after the reset and at the end of the integration period T_F . The difference signal theoretically does not contain the thermal noise component; in practice, however, some amount of the noise is still present.

The most critical problem of insufficient dynamic range, however, is encountered in the detector arrays which operate on the CCD principle. These devices convert the input photon flux into charge which is stored in potential wells and later, after completion of integration, transferred into the charge-to-voltage conversion device. The chief advantage of using CCD arrays in optical signal processing stems from the high packing density of sensing elements and therefore the capability of building large arrays.

The difficulty of extending the dynamic range in these detectors results from the linearity of the photon-to-charge conversion. Amplifiers with nonlinear feedback elements are not easily incorporated into each photosite and the dynamic range extension must be therefore performed directly in the charge domain before the array is read out. We now illustrate the general principle of such a charge domain dynamic range extension. A cross section of the device is shown in Figure 10; it contains a long, narrow photosite with a drain gate on one side. The

drain gate interfaces the photosite with the n^+ drain and allows the charge to be drained out from the photosite. Another gate, not shown in the drawing, runs along the length of the photosite and interfaces it with an element of the readout CCD register. This gate transfers the charge from the photosite to the readout register at the end of the integration period. During the integration period the drain gate is open and the charge is allowed to be continuously drained out. However, due to the relatively long length of the photosite there will be an equilibrium charge distribution along the length, since the new carriers, which are constantly generated at a given rate, cannot be instantly drained out. The equilibrium formation time, which determines the minimum integration time, can be found from the carrier diffusion time along the length of the photosite. The equation describing the charge distribution is as follows:

$$\frac{d}{dx} \left\{ qD \frac{dn}{dx} + \mu \frac{q^2}{c_0} n \frac{dn}{dx} \right\} = -I_s \quad (30)$$

where n represents the charge concentration in electrons per mm^2 , I_s the photocurrent density in amperes per mm^2 and c_0 is the equivalent charge storage capacitance in farads per mm^2 . The rest of the symbols have their usual meaning. At the end of the integration period the drain gate is closed and the charge is quickly transferred into the CCD readout register via the transfer gate. Since the one pixel of the CCD register spans the whole photosite length, the charge transfer time can be much shorter than the integration time which minimizes errors due to the redistribution and integration of unwanted signal. The total amount of the transferred charge can be obtained by integrating equation (30). The boundary conditions are

$$\text{At } x = 0, \quad \frac{dn}{dx} = 0; \quad (31)$$

$$\text{At } x = L, \quad n = 0.$$

The solution for $n(x)$ will be:

$$n(x) = -\frac{kTc_0}{q^2} + \sqrt{\left[\frac{kTc_0}{q^2} \right]^2 + (L^2 - x^2) I_s \frac{c_0}{\mu q^2}}. \quad (32)$$

The total number of electrons transferred will then be

$$N = W \int_0^L n(x) dx, \quad (33)$$

which, after evaluation of the integral, becomes:

$$N = \frac{WL}{2} \left[\frac{kTc_0}{q^2} \right] \left[\left(\frac{a^2+1}{a} \right) \arcsin \left[\frac{a}{\sqrt{a^2+1}} \right] - 1 \right], \quad (34)$$

where W is the width of the photosite. In equation (34) a parameter $a^2 = I_s L^2 \mu / (D^2 c_0)$, where D is the carrier diffusion constant, was introduced to simplify the notation. This result can be further simplified for $a \gg 1$, and the final expression for N rewritten as:

$$N = \frac{\pi}{4} \left(\frac{WL^2}{q} \right) \sqrt{I_s (c_0 / \mu)}. \quad (35)$$

From this result it can be observed that the signal detected at the array is now proportional to the square root of the input photocurrent. This feature doubles the dynamic range. The available CCD well capacity of the transport registers and the charge conversion amplifier capacity will be efficiently utilized. There are other possibilities of charge domain signal processing which can lead to logarithmic or piecewise linear transfer characteristics. The detailed description of such devices, however, is beyond the scope of this paper.

The rate at which A/D's can convert analog signals into a digital format is a strong function of the number of binary bits per sample. Currently, A/D's can digitize a 4 bit sample at 1 GHz rate, 8 bits at 200 MHz, 12 bits at 20 MHz, and 16 bits at about 100 KHz. Precise measurement accuracy (large number of bits) leads to low sample rates, but the machine intelligence community is pushing for higher sample rates. Detection devices which have a logarithmic response to radiation greatly compress the voltage scale presented to the A/D. The first binary bits provide a reasonable measurement accuracy for small signals; the last bits represents huge signal changes and hence poor absolute accuracy.

If measurement accuracy at high levels is not a critical issue, then a single 6 bit A/D converter, sampling at a 200 MHz rate, should be capable of digitizing the outputs of many optical processors. Using a parallel second A/D, in which the

first bit is set at slightly less than the square root of the maximum expected dynamic range, would recover measurement accuracy at high intensities if so desired.

To conclude this section, we stress that there is little difference between the integrating and nonintegrating detector schemes in performance as well as in the possibility of dynamic range extension by utilization of nonlinear elements. The possibility of dynamic range extension in CCD arrays is particularly attractive, since the high pixel packing density and the low noise of charge detection amplifiers can be fully utilized for the large data rates present in optical signal processing applications.

4.2 Data Rate Reduction

The computational power of optical processing results in high output data rates unless steps are taken to introduce on-chip processing that includes nonlinear or decision operations. We discuss three broad possible methods to achieve the desired data rate reduction: (1) on-chip processing for arrays that output electronic signals, (2) methods for segmenting arrays, along with the possibility of on-chip processing, and (3) new array types in which the output of the detector remains in an optical format for further processing.

4.2.1 Optical-In, Electrical-Out Arrays

Signal processing within the detector chips may significantly reduce the output data rates to the postprocessor. Some of the desired functions are:

- o Video amplitude compression (perhaps programmable)
- o Temporal change detection from frame-to-frame
- o Spatial change detection along the elements of the array
- o Dynamically programmable spatial convolutions for such tasks as centroiding or corner detection, using cellular blocks of up to 7x7 elements
- o Random access to any subregion of the array to isolate and dynamically track selected activity
- o Threshold levels that may be either globally or locally set and adjusted adaptively to achieve constant false alarm rates
- o Methods for synchronous detection to remove the strong bias terms or background that arise in some processing operations
- o Analog first-in, first-out (FIFO) or last-in, first-out (LIFO) memory

While some of these functions have already been demonstrated or designed at low or kilohertz rates, the tremendous data rates produced in some real-time processing operations require frame rates that are two orders of magnitude faster (e.g., 100MHz to 1 GHz), which implies that the various analog operations cited must be performed in 10 ns to 1 ns at each element in the array.

Such a tremendous leap in capabilities is likely to require new materials or new techniques such as quantum well structures. Some of the more critical difficulties involve readout schemes with contradictory demands for large bandwidth, low noise, and low power, commensurate with the anticipated load. Important characteristics such as quantum efficiency, gain, offset, and threshold must be uniform both within each chip and from chip to chip. To improve performance, it may be useful to provide a three-dimensional photodetector structure so that the entire pixel area can be used for imaging the light signal; the complex electronic processing and multiplexing circuitry needed to perform the on-chip operations would then be placed on the opposite side of the chip. A first implementation may be the use of backside illumination with the photosignal coupled to signal processing circuitry on the front side. A more general and longer range implementation would be the use of multiple silicon-on-insulator layers with the photodetector being formed in the top layer and the circuitry being implemented in the intermediate layer and substrate. It is important that the photodetector imaging layer be compatible with the laser source. Thus, if the laser wavelength is $\approx 850\text{nm}$, then either a very thick silicon imaging layer must be used to obtain sufficient quantum efficiency or some other semiconductor with a high absorption coefficient must be used.

An important family of optical processors operate in a heterodyne fashion to increase the useful dynamic range of the system or to measure both the amplitude and phase of the light distribution. In either case the IF frequency may range from 1MHz to 30 MHz. For these applications, the detector array must be nonintegrating and the readout structure is generally a set of parallel outputs. An acousto-optic channelizer typically produces 100-1000 narrowband channels, each centered at an IF frequency. By synchronously demodulating each IF channel on the photodetector chip, we can obtain the RF envelope of each channel at baseband which is more suitable for other channel processing/detection functions on the photodetector chip.

In some applications, we would prefer to operate on signals that have not

been downconverted from their original RF band. A programmable RF photodetector could sum the RF photodetector currents from a (typically) linear array to a common RF output, possibly with controlled summing gain at each photodetector element. By placing this RF photodetector in the frequency plane of a coherent optical processor, we can implement programmable filter functions such as frequency excision (notch rejection filters), tunable high-Q bandpass filters, and filter equalization functions. A high contrast (40-80 dB) on/off ratio for each photodetector pixel is desired to perform various matched filter functions.

4.2.2 Segmented Arrays

Nature has designed a number of imaging systems with specially designed detector planes that reduce the data rates that must be processed. An example is the eye, where only the central viewing zone, called the fovea, has high resolution. The cone detectors, located in the fovea, provide color vision while the rod detectors, located almost exclusively outside the fovea, have no color sensitivity. The image plane is nonuniformly sampled and many detectors (e.g., the rods) are interconnected at a number of levels. Thus the detector plane of a vertebrate is arranged to reduce the data rate transmitted to the brain.

We wish to explore the possibility of interconnecting the detector plane of an optical processor so that we may reduce the data rate of the signals provided by the optical processor. In general, we expect to find that the requirements placed upon the detector array for the optical processor will differ from those of an imaging system. One attempt to design a special detector for an optical processor resulted in the "ring-wedge" detector discussed in Section 2. The detector, made on a single silicon wafer, has a sensitivity of 0.25 amps/watt at 633 nm and has found application in the on-line inspection of products such as hypodermic needles.

Some optical processors use conventional detectors but modify the illumination at the detector plane by using a mask. Size distribution measurements can be made by placing a mask in the back focal plane of a lens and by illuminating a medium containing scattering objects whose size distribution is of interest. At least two commercial devices make use of this approach²⁴. In the following discussion we wish to generalize the concept of specially designed detector systems for optical processing.

Some systems, such as those using a quadrant detector for tracking, use a feedback system to bring the optical signal to the optimum position on the detector. This is similar to the operation of the eye where the image of interest is brought to the part of the detector plane where the high resolution elements are located. More generally, an array with a random access capability allows the user to define a small window of active detector elements and to move that window around on the focal plane. If an important event is known to exist in a given region, that region can be sampled at a faster rate while letting regions of lower interest integrate for longer periods.

One of the easiest modifications of a detector array to accomplish is segmenting the array of N elements into subarrays. Each subarray is given the capability of processing the signal at a local level to reduce the amount of data it will report to the optical processor output. If we have K subarrays, each of size M , the total number of detectors we must scan if there had been no subdivision is $N=KM$ and the data rate would be

$$\frac{NQ}{t} = \frac{KQM}{t}, \quad (36)$$

where Q is the number of quantization levels of the signal and τ is the system clock speed. The assumption is that the system is limited to a data rate of

$$\frac{dr}{dt} = \frac{MQ}{\tau}. \quad (37)$$

By dividing the detector array into K subarrays we reduce the data rate to the desired level. In each subarray, some processing is done to simplify the amount of data by S , where $S < 1$, so that each subarray provides $B = SM$ data points. A time T is spent performing this processing so that we cannot scan the subarrays at a rate faster than K/T , where $T = I\tau$ and I can be thought of as an effective integration factor, $I > 1$. The result of the processing at the subarray level is to reduce the total data rate to

$$\frac{dR}{dt} = \frac{KBQ}{T} = \frac{KSMQ}{I\tau} = \frac{KS}{I} \frac{dr}{dt}. \quad (28)$$

Thus processing at the subarray level, characterized by the ratio S/I , can have a major impact on the processing rate needed at the post detection stage of the system.

An example may help support this simplified analysis of the problem. First, let us assume that we use a one dimensional array of detectors in an optical

spectrum analyzer. Suppose that we have a linear array of 4000 detectors. If 60 signals were encountered during a frame time of 1 ms, we would need a data rate of 81.92 megabits/sec, using a 20-bit word to achieve the desired accuracy. Suppose now that we divide the array into 32 subarrays of 125 detectors each. We also assume that we will not encounter more than two signals in each subarray so the chance of saturating our subarray is small.

Each subarray will report out the position and size of each signal it encounters. We digitize the signal according to the following rule

5 bits → label for subarray

7 bits → label for position in subarray

20 bits → magnitude of signal (60 db dynamic range)

Following these rules, the total bit rate for the same 60 signals is only 2.56 megabits/sec, a reduction in the data rate of 42 times. The identification of the pixel within the subarray can be done with a simple counter and would not measurably increase the data rate of the subarray processor. The label of each subarray does not change and would involve no processing.

Another optical processing problem is the identification and location of an object in the field of view of a two dimensional optical processor. This problem is usually solved by performing a correlation operation, as we discussed earlier. For most complicated signals, the correlation function is nearly a delta function and we need only locate the position of maximum light intensity. To measure the position of a correlation peak, we need to determine the intensity and the position of the correlation signal. For output signals of this type, a detector based upon the lateral photoeffect can be used²⁵. The device has a floating p-n junction and the photovoltage is measured by contacts placed on either side of the junction. A spot of light illuminating the center of the junction produces no voltage but a spot of light close to one of the contacts produces a voltage proportional to the distance from the center of the junction. The polarity of the photovoltage depends on which side of the center the light spot is found.

The lateral photoeffect detectors are usually designed as a linear or circular detector. Other configurations could be envisioned such as a spiral or a zigzag configuration. More important than the configuration of the detector is the method used where multiple signals may produce multiple correlation peaks. Each signal will produce a correlation peak located at a position corresponding to the position of the signal in the input field of view. Segmentation of the

output plane may also be a solution to this problem. Each segment would contain a lateral photoeffect detector. The size of the detector would be dictated by the expected size of the pattern and the probability of finding two patterns within a given region.

4.2.3 Optical In/Optical Out Focal Plane Detectors

Some optical processors are capable of producing outputs with two spatial dimensions and one temporal dimension which may require spatial sampling with 1000x1000 samples and temporal sampling at rates approaching 1GHz. No practical sensor is foreseen which would be capable of bringing out all of the data electronically. A possible solution is to develop a detector in which the output is in an optical form to provide the required output bandwidth and parallelism. The optical output detector is envisioned as a two-dimensional array of elements which respond to light, provide gain, perform additional operations (preferably nonlinear) in response to some external control signal, and retain the signal in an optical form. The optical output provides the opportunity for additional optical processor stages to follow. The optical output may be coherent or incoherent; it could even be of a different wavelength from the input. We see a broad applicability for such devices because they represent a large variety of 3-terminal nonlinear optical devices. The avenues of approach to such devices may include research into GaAs/GaAlAs superlattice materials.

5. Summary and Conclusions

Photodetector arrays have been, and continue to be, designed primarily for imaging applications. Optical signal processing applications place new and more stringent requirements on these devices. We need less blooming, faster recovery, less electrical readout noise, more video output lines, segmented arrays, and so forth. The major improvements, however, must be in dynamic range and in techniques for reducing the output data rates.

We have briefly reviewed typical optical processing applications, described the basic physics of the devices, and have offered some suggestions for improving the dynamic range and for reducing the output data rates. It is our hope that the ideas presented here will result in the development of devices that are better tailored to the needs of optical computing systems.

This paper was the result of an Army Research Office Palantir study. These studies address the physical foundations of approaches to solutions of important

technological problems with the aim of stimulating new avenues for progress toward their solution. The participants in the study consisted of the authors of this paper. One of the authors (G.W.A) wishes to acknowledge partial support by the Department of the Navy and (R.J.K) the partial support by the Department of the Air Force.

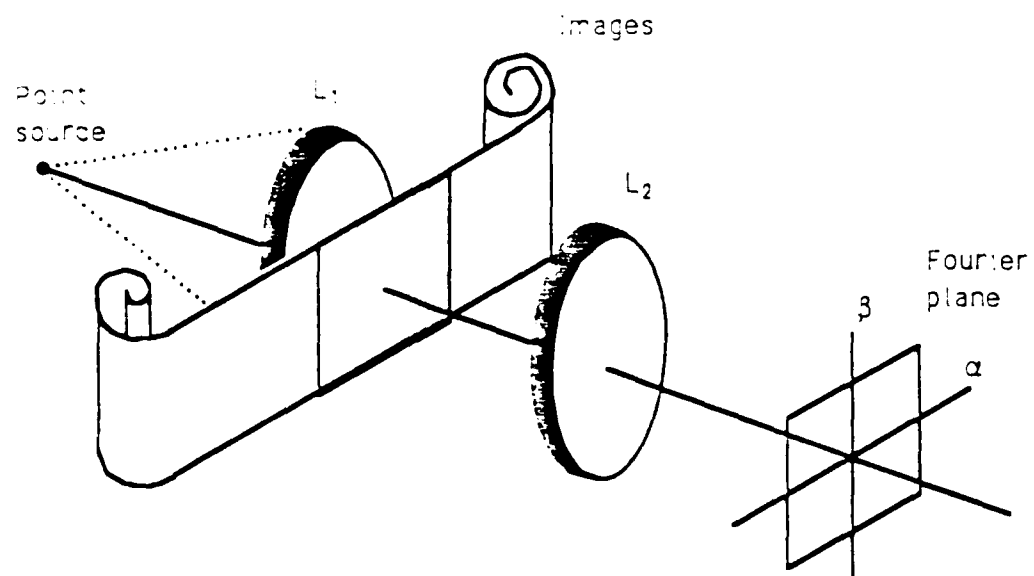
References

1. A. VanderLugt, "Coherent Optical Processing," Proc. IEEE, Vol. 62, p. 1300 (1974).
2. J. W. Goodman, "Operation Achievable with Coherent Optical Information Processing Systems," Proc. IEEE, Vol. 65, p. 39 (1977).
3. Special Issue on Optical Computing, Proc. IEEE., Vol. 65, January 1977.
4. Special Section on Acousto-optic Signal Processing, Proc. IEEE, Vol. 69, January 1981.
5. Special Issue on Optical Computing, Proc. IEEE, Vol. 72, July 1984.
6. Special Section on Optical Pattern Recognition, Optical Engineering, Vol. 23, Nov./Dec. 1984.
7. A. VanderLugt, "Operational Notation for the Analysis and Synthesis of Optical Data Processing Systems," Proc. IEEE, Vol. 54, p. 1055 (1966).
8. N. George, J. Thomasson, and A. Spindel, 1970, "Photodetector for Real Time Pattern Recognition", U. S. Patent No. 3689 772
9. J. T. Thomasson, T. J. Middleton, and N. Jensen, SPIE Proc. Coherent Optics in Mapping, Vol. 45, p. 257 (1974).
10. G. Lukes, SPIE Proc. Coherent Optics in Mapping, Vol. 45, p. 265 (1974).
11. A. Korpel, "Acousto-optics -- A Review of Fundamentals," Proc. IEEE, Vol. 69, p. 48 (1981).
12. E. H. Young, Jr. and S-K Yao, "Design Considerations for Acousto-optics Devices," Proc. IEEE, vol. 69, p. 54 (1981).
13. R. Doyle and W. Glenn, "Remote Real-time Reconstruction of Holograms Using the Lumatron," Appl. Opt., Vol. 11, p. 1261 (1972).
14. "Real-time Modulator for Coherent Optical Processing," Final Report AFAL-TR-73-88, Environmental Institute of Michigan, May 1973.
15. See, for example, W. T. Rhodes, "Acousto-optic Signal Processing: Convolution and Correlation," Proc. IEEE, Vol. 69, p. 65 (1981).
16. T. M. Turpin, "Spectrum Analysis Using Optical Processing, Proc. IEEE, Vol. 69, p. 79 (1981).

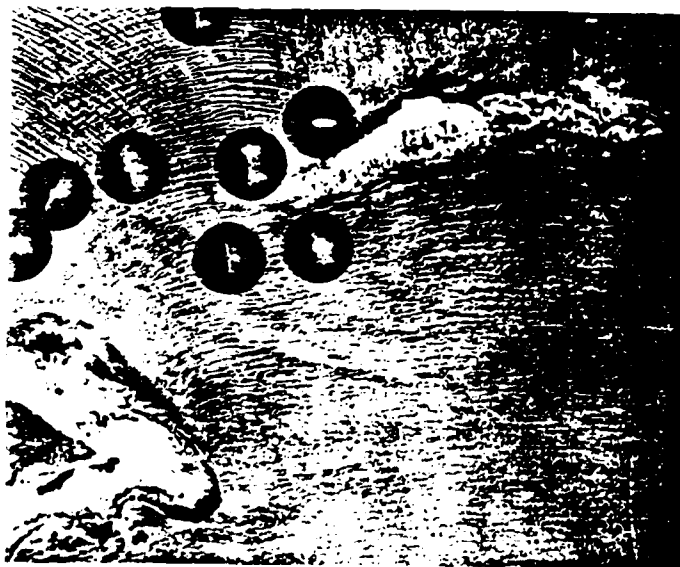
17. A. VanderLugt, "Interferometric Spectrum Analyzer," Appl. Opt., Vol. 20, p. 2770 (1981).
18. A. VanderLugt and A. M. Bardos, "Spatial and Temporal Spectra of Periodic Functions for Spectrum Analysis," Appl. Opt., Vol. 23, p. 4269 (1984).
19. D. F. Barbe, "Imaging Devices Using the Charge-Coupled Concept", Proc. IEEE, Vol. 63, p. 38 (1975)
20. G. M. Borsuk, "Photodetectors for Acousto-optic Signal Processors," Proc. IEEE, Vol. 69, p. 100 (1981).
21. G. M. Borsuk, G. W. Anderson, and F. J. Kub, "Photodetectors for Acousto-optic Signal Processing", SPIE, Vol. 639, p. 2, (1986)
22. S. G. Chamberlain and J. P. Y. Lee, "A Novel Wide Dynamic Range Silicon Photodetector and Linear Imaging Array", IEEE Trans. Elec. Devices, Vol. ED-31, p.175, (1984)
23. M.H. White, D.R. Lampe, F. C. Blaha, and I. A. Mack, "Characterization of Surface Channel CCD Image Arrays at Low Light Levels", IEEE J. of Solid-State Circuits, Vol. SC-9, p. 1, (1974)
24. MICROTRAC, Leeds and Northrup, North Wales, Penn. and Particle and Droplet-Size Distribution Analyzer, Type ST 1800, Malvern Instruments, Malvern, England
25. J. T. Wallmark, Proc. IRE 45, 474 (1957) and J. I. Alferov, V. M. Andreev, E. L. Portnoi and I I. Protasov, Sov. Phys. Semiconductors 3, 1103 (1970)]

Figure Caption List

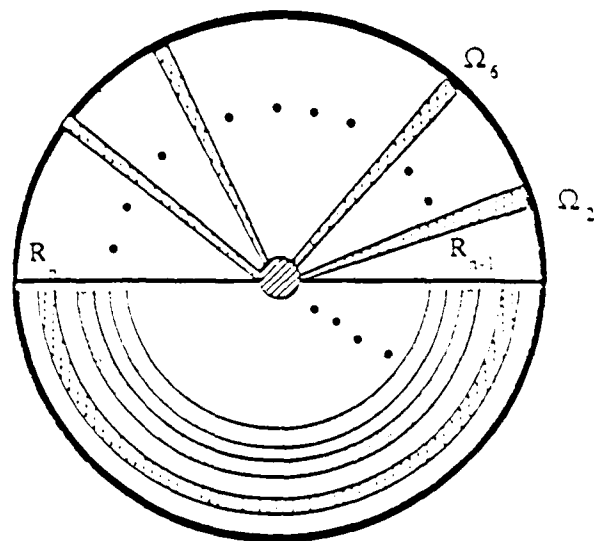
- Figure 1 Spectrum Analyses
- 1a) Optical Spectrum Analyzer
 - 1b) Spectra of Wave Patterns
 - 1c) Ring/Wedge Detector Geometry
- Figure 2 Real-time Spectrum Analyzer
- 2a) The Bragg Cell Spectrum Analyzer
 - 2b) The Geometry of the Desired Array
- Figure 3 Heterodyne Spectrum Analysis
- 3a) Interferometric Optical System
 - 3b) Detector Circuitry (RF Receiver)
- Figure 4 Correlation System for Pattern Recognition
- Figure 5 Pattern Recognition
- 5a) Input Scene
 - 5b) Correlation Plane
 - 5c) Scan Through Correlation Peak
 - 5d) Desired Nonlinearity
- Figure 6 PN Junction Used as a Photodiode
- Figure 7 Crosstalk/Depletion Region
- Figure 8 Feedback Amplifier
- Figure 9 Integrating Detector with Feedback
- Figure 10 Photodetector with Drain Gate



(a)



(b)



(c)

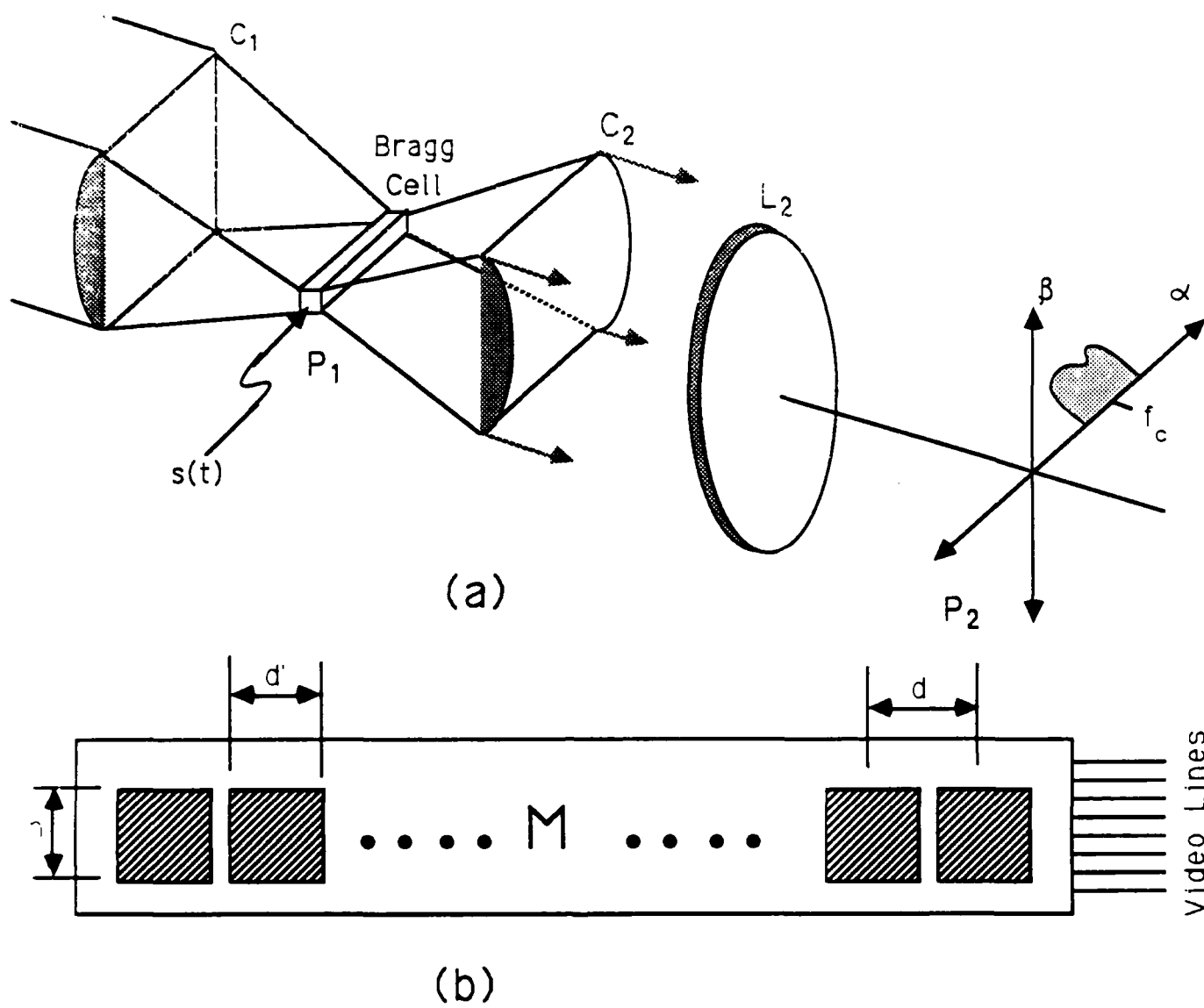


Figure 2

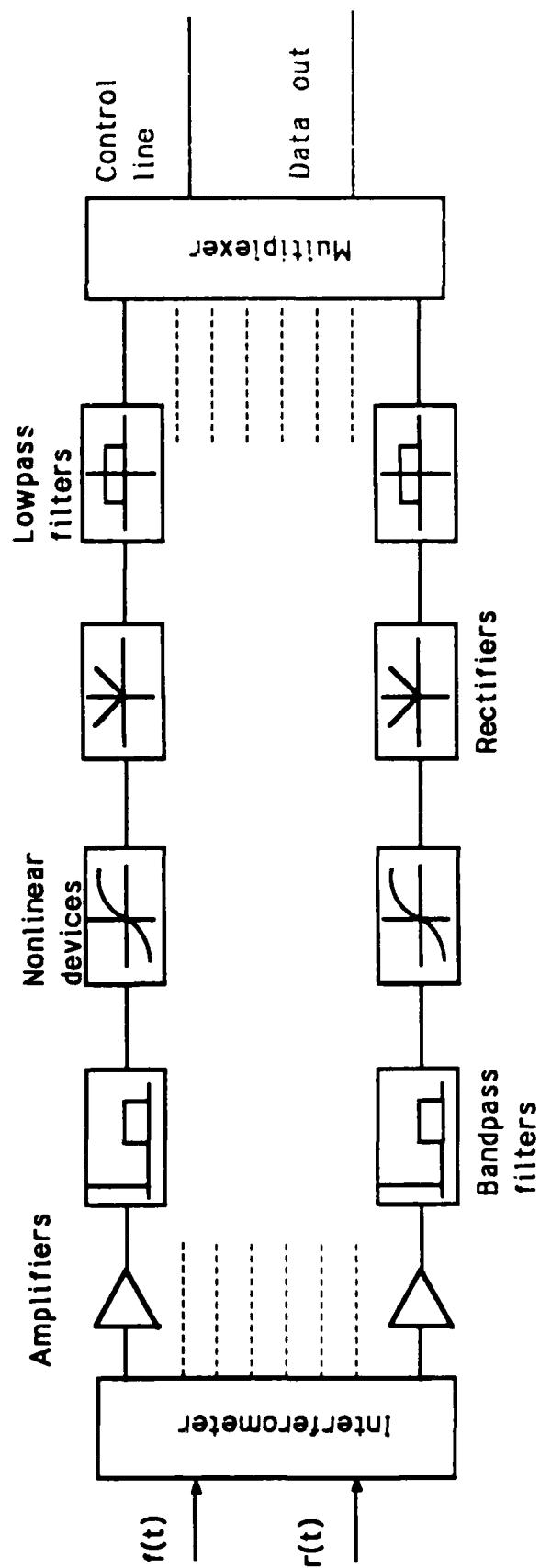
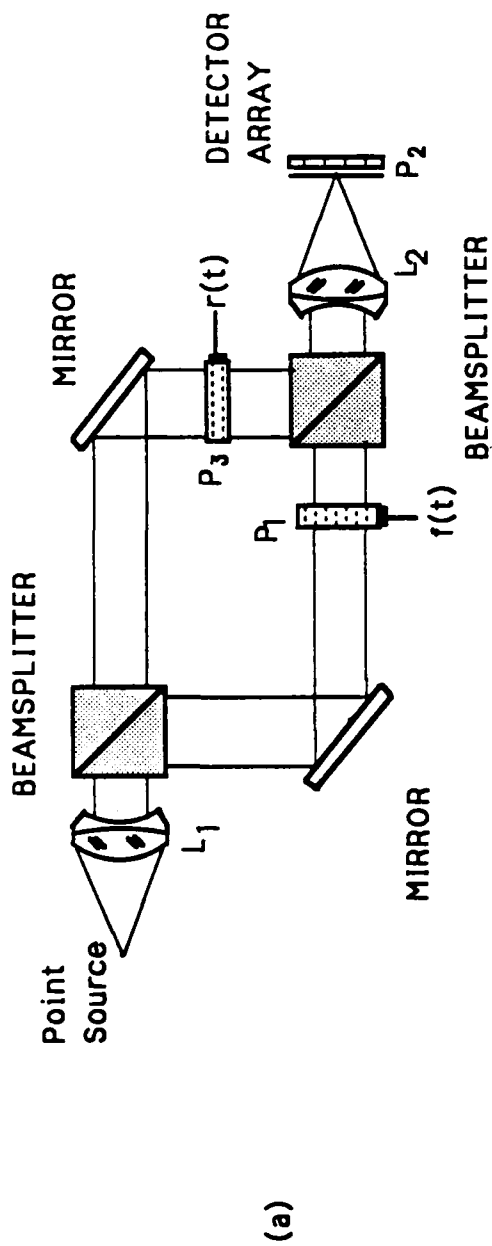


Figure 3

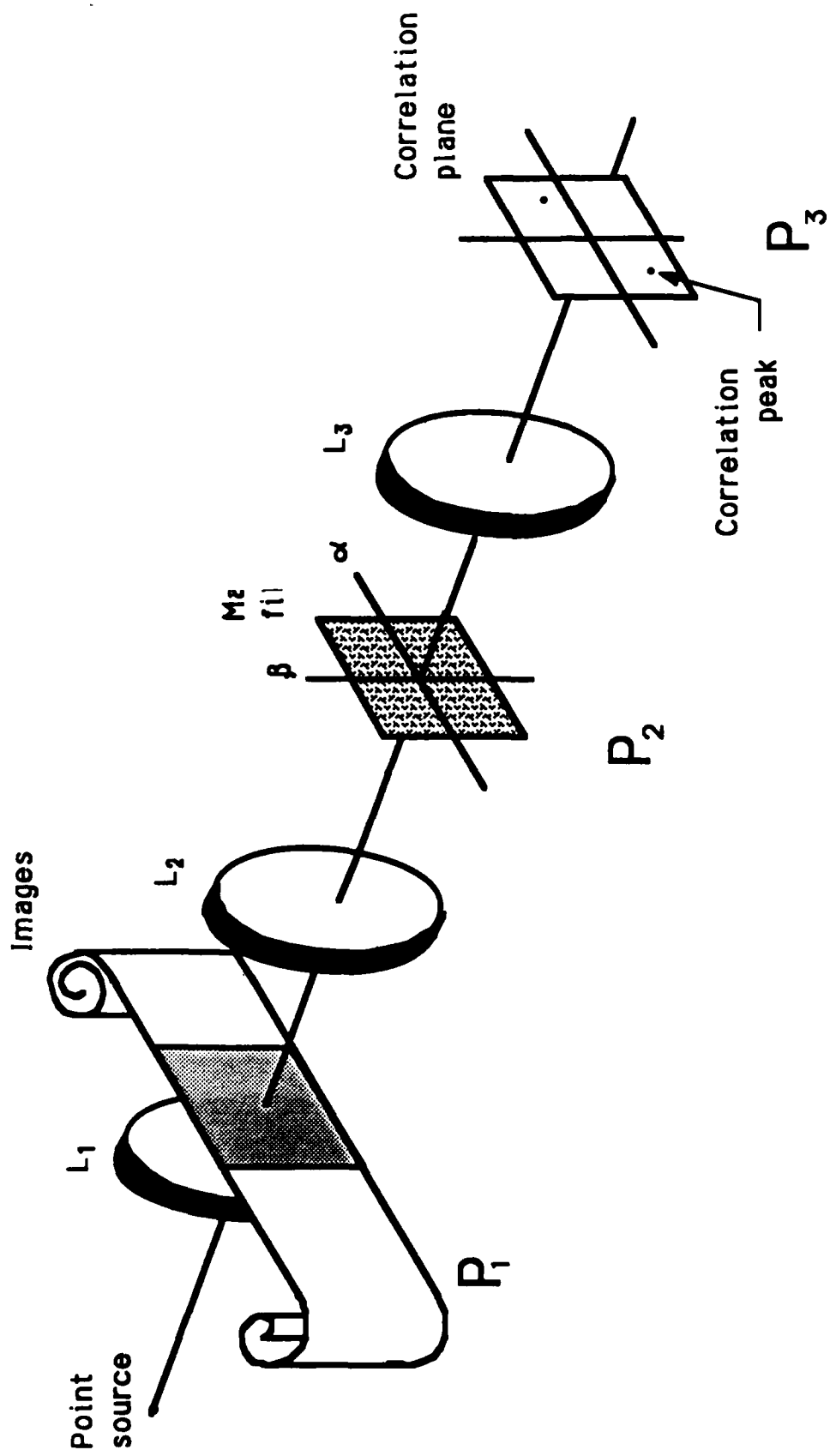
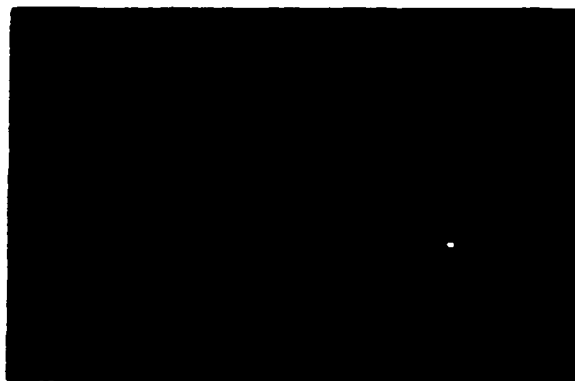


Figure 4

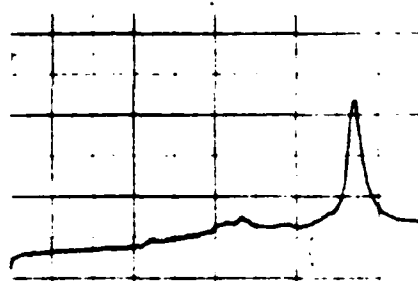
(a)



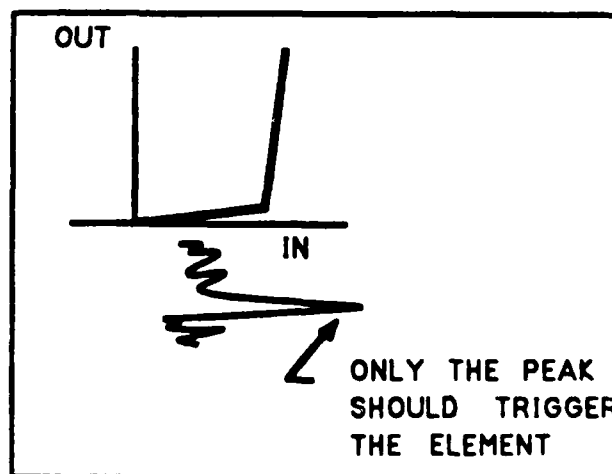
(b)



(c)



(d)



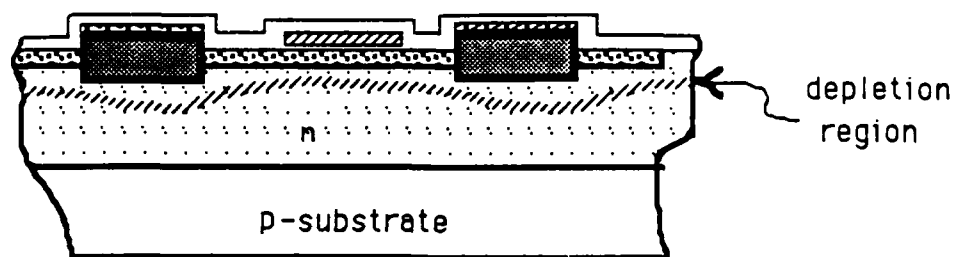


Figure 6

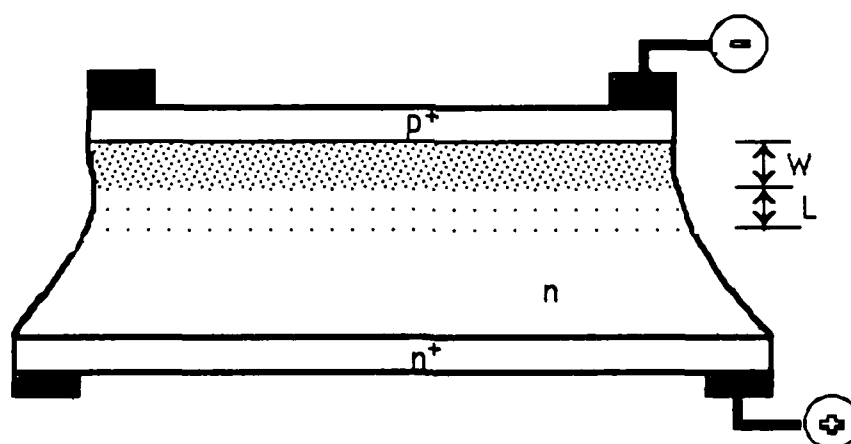


Figure 7

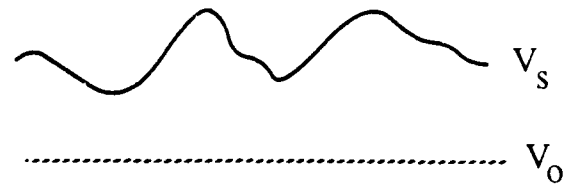
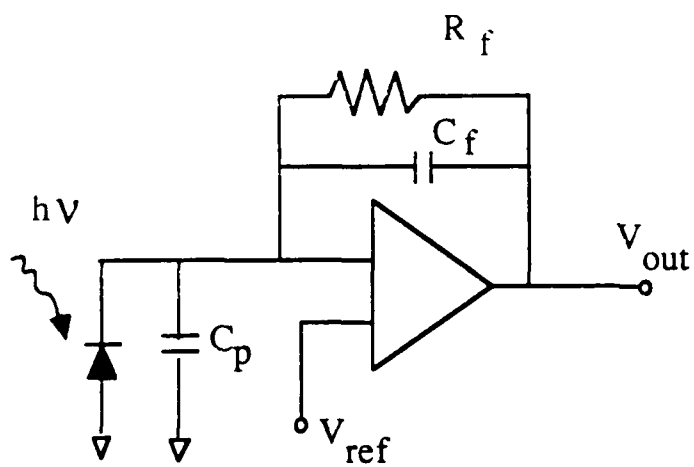


Figure 8

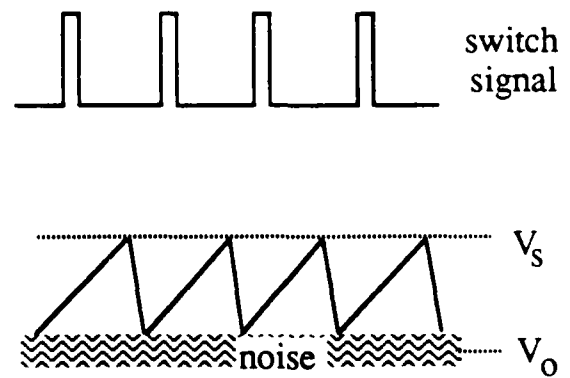
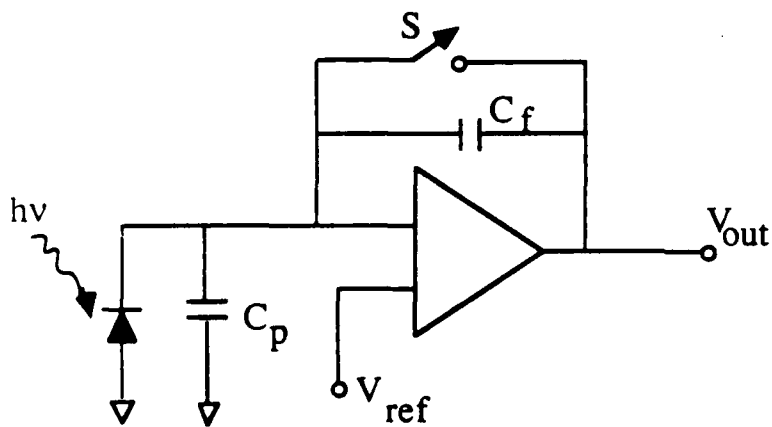


Figure 9

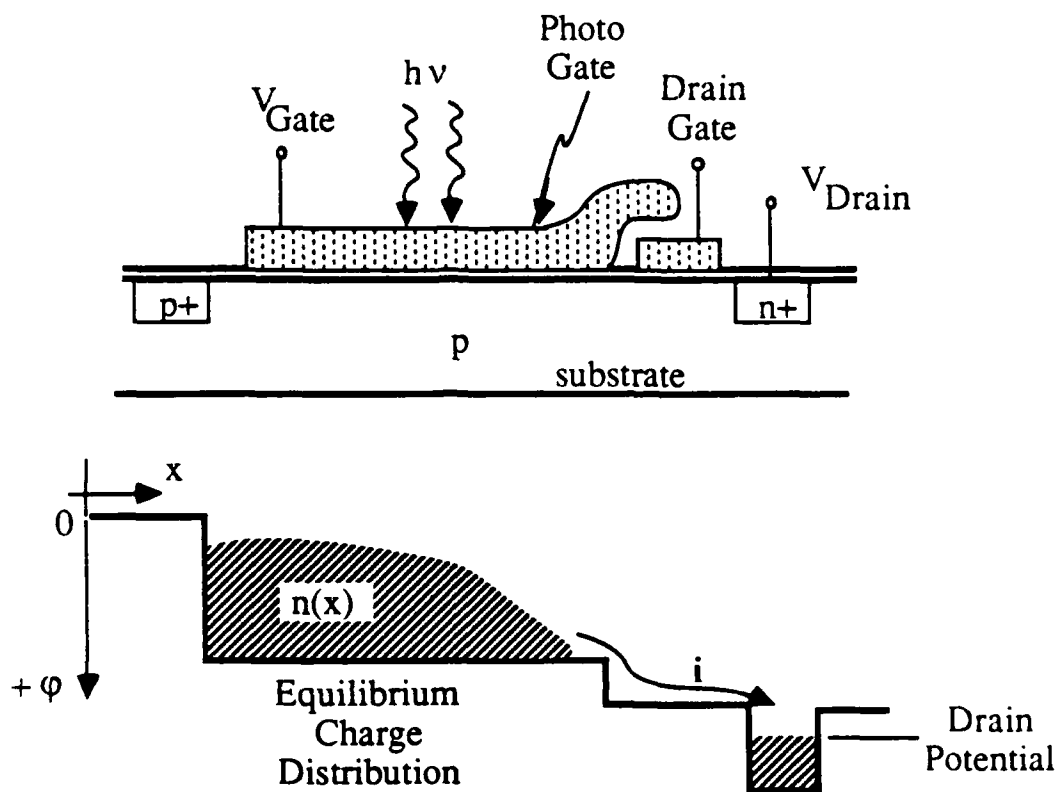


Figure 10

END

FEB.

1988

DTic